Reply to the comments of Referee 2 by H.E. Rieder on behalf of the authors.

We thank the referee (called "R2" below) for the discussion and like to clarify several points. Unfortunately, he/she remains unconvinced:

R2: "However, the data presented in this supplement and the interactive comments do not change my opinion that thresholds for ELOs and EHOs depend strongly enough on the selected reference period to invalidate the conclusions drawn in the study … one important time period is still missing, ie. the one with low ozone: 1990-2008 …"

Although we agree that the selection of the periods for determination of the thresholds is relevant for the study, we would like to emphasize that the results most relevant for the main conclusions are not the thresholds themselves but their effects on the time series of the frequencies of the EHOs and the ELOs. In Fig. 1 below we show that the frequency distribution of ELOs and EHOs and therefore fingerprint detection is robust also if thresholds for ELOs and EHOs are chosen from the time period 1989-2008 (please note that we add 1989 to the time interval suggested by R2 to be consistent with the other 20 year periods shown in the supplement).

In his/her reply of 13 July 2010, the referee introduces the time period 1990-2008 (which differs from those suggested in his/her original referee comment) to characterize the "low ozone period" (1992-2008). To address the comment of R2 we include this time period in the analysis where thresholds for ELOs and EHOs are estimated also on the time period 1989-2008.

Figure 2 shows the ratios between the differences in thresholds for extremes and mean values derived for the periods 1980-99 minus 1960-2008 and 1989-2008 minus 1960-2008. With a few exceptions (particularly concerning the October values) it is obvious that the difference among mean values is significantly larger than those between threshold values (on average by about a factor of 2).
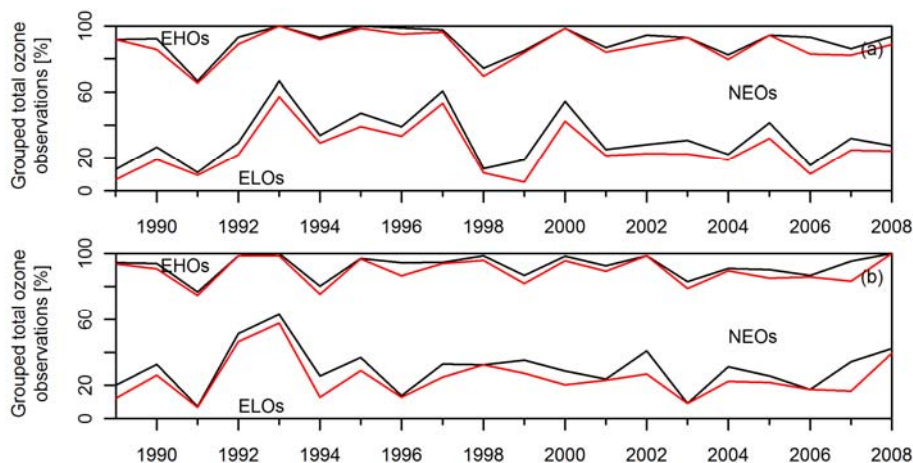

Figure 1: Evolution of fractions of days with extreme low (ELOs) and high (EHOs) total ozone and not extreme days (NEOs) for (a) spring (MAM) and (b) winter (DJF) during the period 1989-2008 based on different thresholds applied. Black lines: based on thresholds estimated from the time period 1960-2008 (used in the analysis of Rieder et al., 2010a,b). Red lines: based on thresholds estimated from the time period 1989-2008 (i.e., one of the periods suggested by the reviewer).
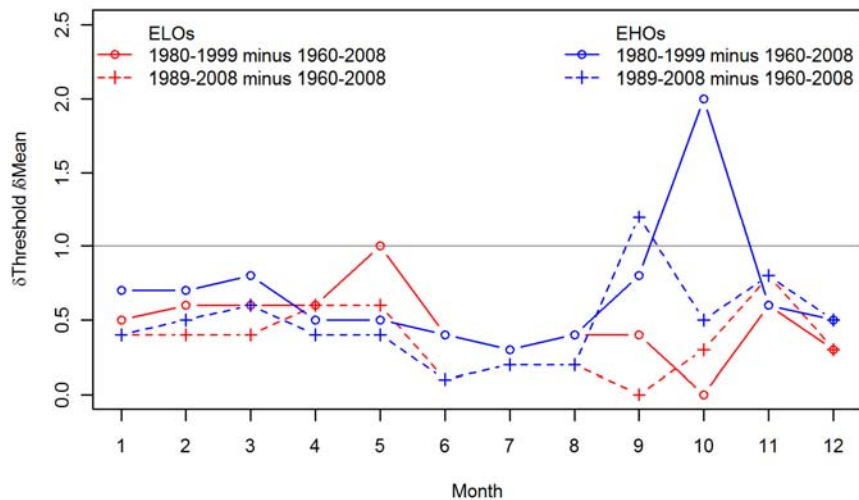
Figure 2: Ratio of differences (δ) in threshold values for extremes for different time periods over their difference in mean values. The latter is typically larger by a factor of ~2, irrespective of the time periods chosen (long-dashed: 1980-1999 minus 1960-2008; short-dashed: 1989-2008 minus 1960-2008) or the type of extremes (red: ELO; blue: EHOs).

The analysis and the additional figures presented here will be included in the updated supplementary material. They underline that threshold selection within the perturbed time period does not affect the shape of the frequency distribution of the extremes in any significant way, and therefore neither the fingerprint analysis nor the conclusions drawn in our papers.

Furthermore the Referee states:

R2: "Referring again to page 12776, lines 16-19, of the manuscript, I rewrite the statement: "This shows that thresholds for ELOs and EHOs depend (only) by a factor of 0.7 compared to the mean values on the selected period, so changes in the frequency of ELOs and EHOs cannot be attributed to a particular choice of reference period." A factor of 0.7 is not small enough to justify the conclusion in the second half of the sentence! The authors argue that the same message ("weakly" dependence) is visible in the data presented in the supplement of the manuscript, ie. Figs. S2 and S3. In Fig. S2 I suppose to see similar factors as stated above."

We respectfully disagree, but to ensure clarity we will rewrite the statement on P12776 as follows (with Figure 2 from this reply becoming Figure S5 in the updated supplement):

"The monthly mean values of the time series (in Figure 8 and Figure S5) show larger differences than the corresponding thresholds for low and high total ozone. Changes in the frequency of ELOs and EHOs within the perturbed time period cannot be attributed to a particular choice of the reference period, as threshold estimates depend only weakly on the reference period chosen (see also extensive "Interactive Discussion" to this article). Rather, such changes must be interpreted as fingerprints of the physical factors affecting ozone."

Finally the Referee states:

R2: "The authors put forward good reasons not to detrend their data set before further analyses and not to use different sub periods. However, this does not allow to draw automatically the converse argument that valid results comes out of a study without

detrending or using less sub periods. Sometimes one has to admit that a data set is not good enough to retrieve the desired conclusions. In my opinion, this is here the case, with so much decadal variability, which cannot be addressed adequately by using the selected time periods."

We have to admit, that we do not really understand this statement and therefore are unable to reply in any appropriate way. Nevertheless we hope that the reply and additional figures help answering the questions asked by R2 and strengthen the value of our analysis.